# Array-Ready Oligo Set™ for the Human Genome Oligo Set Version 1.1

Our Human Genome Oligo Set Version 1.1 contains 16,659 70mer probes. This set contains an upgrade set of 2688 probes plus the original 13,971 Version 1 probes. An amino linker is attached to the 5' end of each oligo.

Gene sequence source and selection

The original 13,971 probes are designed from the UniGene Database Build Hs 119 (July 2000) developed and maintained at the National Center of Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov). The upgrade set was designed from UniGene Database Build Hs 133 (April 2001).

| Number of oligos | Database |
|---|---|
| 13,971 | Build Hs 119 |
| 2688 | Build Hs 133 |
| 16,659 | Total number of oligos |

*Advantages of using gene sequences from UniGene*

UniGene is an open source database freely available to everyone. UniGene automatically clusters GenBank sequences into a nonredundant set of gene-oriented clusters. It has become one of the most widely used *de facto* standards in the public domain for cataloguing human genes. Each UniGene cluster contains cloned genes and expressed sequence tag (EST) sequences that represent a unique gene. UniGene sequences are filtered for contaminant sequences, genomic repetitive regions, and low-complexity sequences using NCBI's Dust program. All oligos are designed from the representative sequence of each of the 16,659 clusters. This chosen representative sequence is the sequence with the longest region of high-quality sequence in each cluster.

The 16,659 70mer probes represent 16,257 genes when mapped to UniGene Build Hs 133. The number of genes is less then the number of oligos due to a number of factors including merging of UniGene clusters or sequences.

## Probe design

Once a gene has been selected to be included in the set, a probe is selected with an optimal set of parameters. Large numbers of 70mer candidate probes for each gene are selected using the following criteria for the Human Genome Oligo Set. All oligos are within 78°C ±5°C. using the following formula:

$$T_m = 81.5 + 16.6 \text{ x } \log[Na+] + 41 \text{ x } (\#G + \#C)/\text{length} - 500/\text{length}$$
$$\text{where } [Na+] = 0.1 \text{ M and length} = \#A + \#C + \#G + \#T$$

Each oligo is within 750 bases from the 3' end of the available gene sequence and is scanned for potential hairpins. Once an oligo candidate has been selected within the criteria listed above. For all oligos in the Human Genome Oligo Set Version 1.1, using BLAST, each oligo is aligned against all representative sequences in the UniGene database used to design the original sequence. Using the alignment with the candidate oligo versus the highest scoring non-self gene, a BLAST percent identity score is computed. The highest scoring non-self gene is defined as the sequence that yields the most matched bases in an alignment. This BLAST percent identity is also referred to as cross-hybridization identity of the oligo.

This calculated percent identity score is dependent on the size of the sequence database used to BLAST against, oligo sequence, and use of either gapped or no-gap alignment method.

One final oligo for each gene is selected with the minimum percent identity or cross-hybridization identity.

Please note for a small number of genes that did not yield oligos satisfying all the above criteria certain rules were relaxed. For those genes, the oligo is selected anywhere in its sequence or is designed to be less or greater than 70 bases long.

Due to high conservation of sequence or percent similarity between a small number of genes, 112 oligos in the set have 100% identity over the length of oligo to another oligo in the set.


## Design criteria distributions

The following illustrations show the distribution of all 16,659 oligos for melting temperature, GC content, location from 3' end of gene sequence, length of maximum hairpin stem length, and BLAST percent identity or cross-hybridization similarity.

In Figure 5, to generate the cross-hybridization identity graph, each oligo was BLASTed against all the representative sequences in UniGene Build Hs 133.
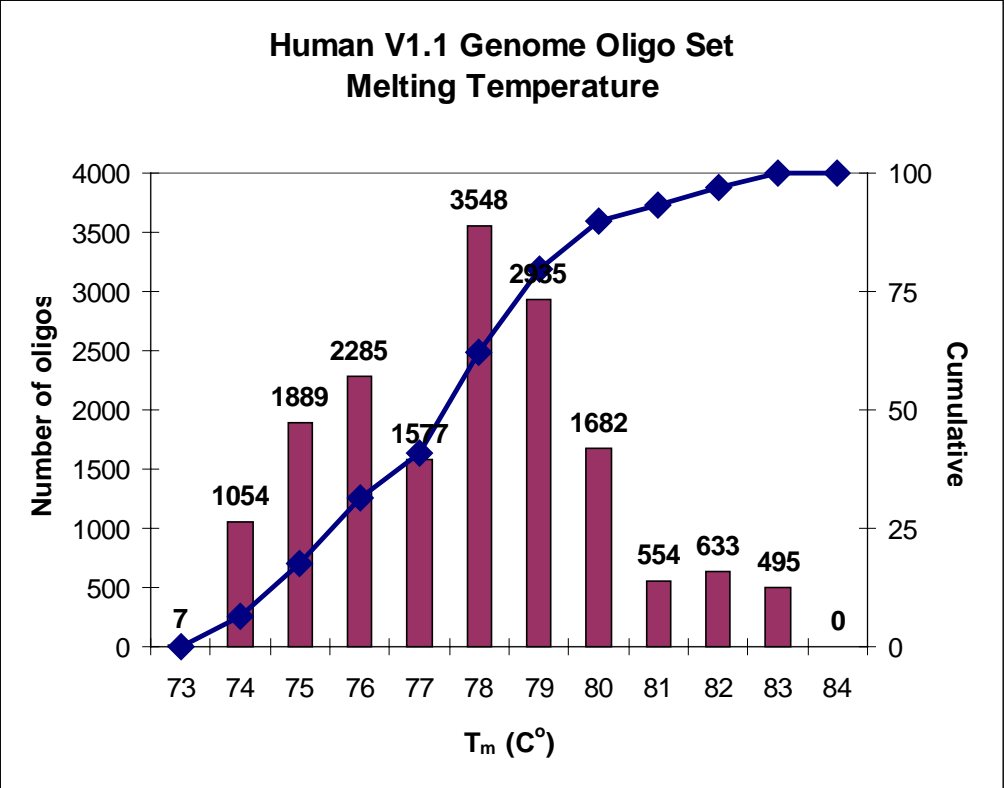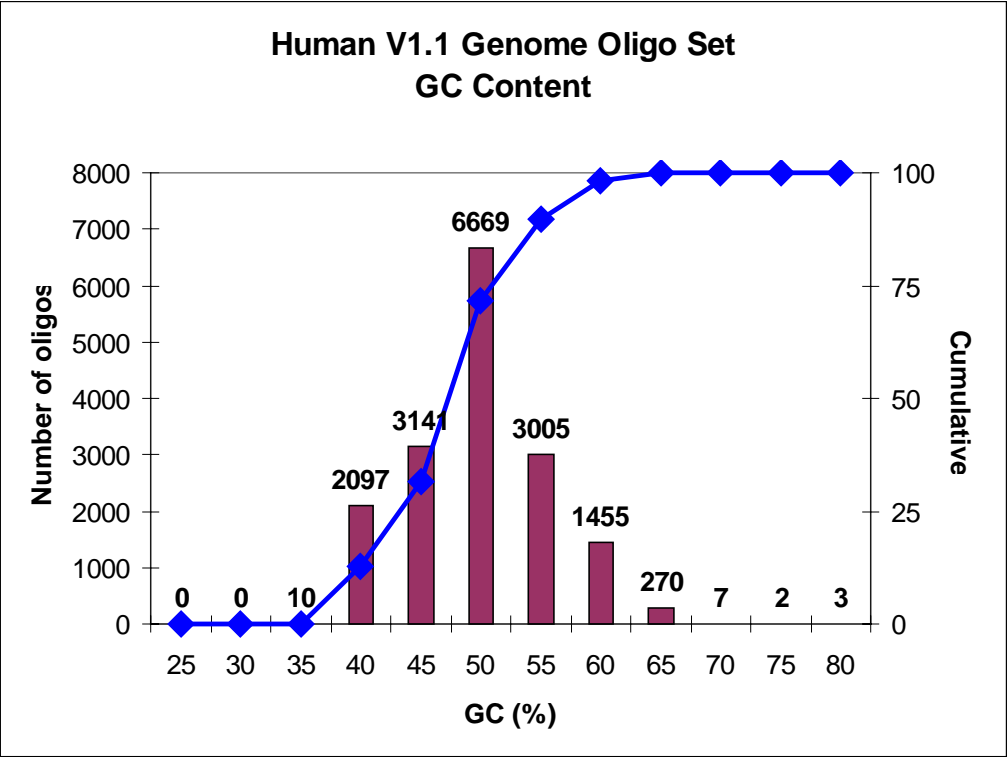
Figure 1. Melting Temperature



Human V1.1 Genome Oligo Set
Melting Temperature

Figure 2. GC Content



Figure 3. Location from 3' End
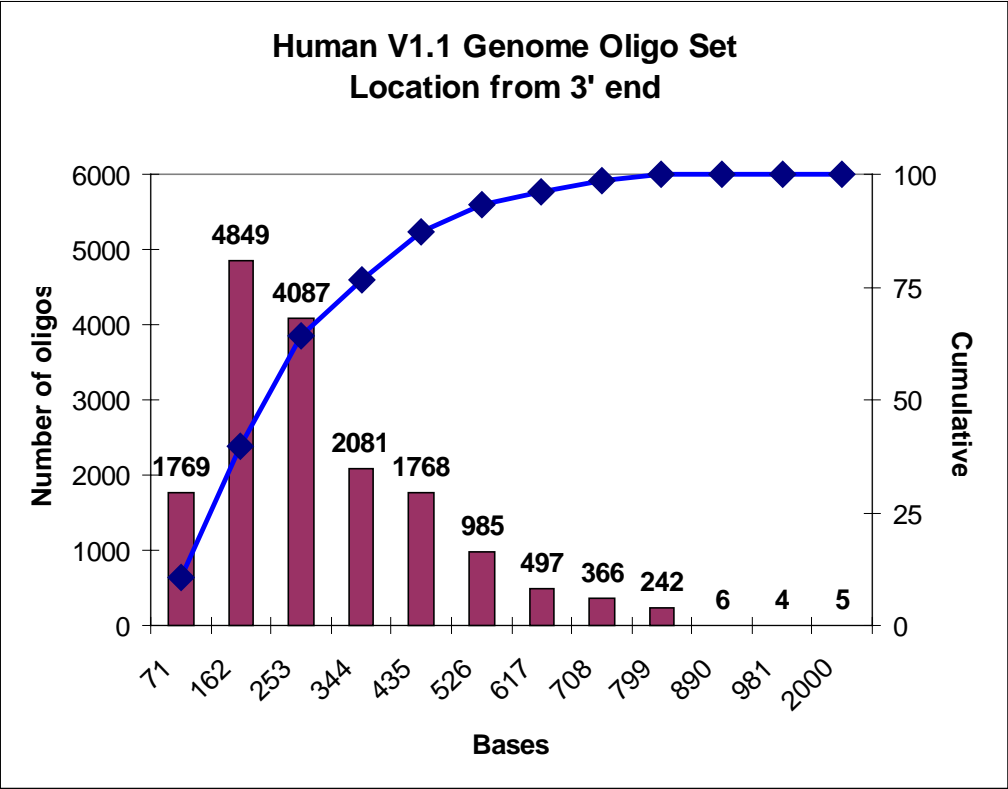
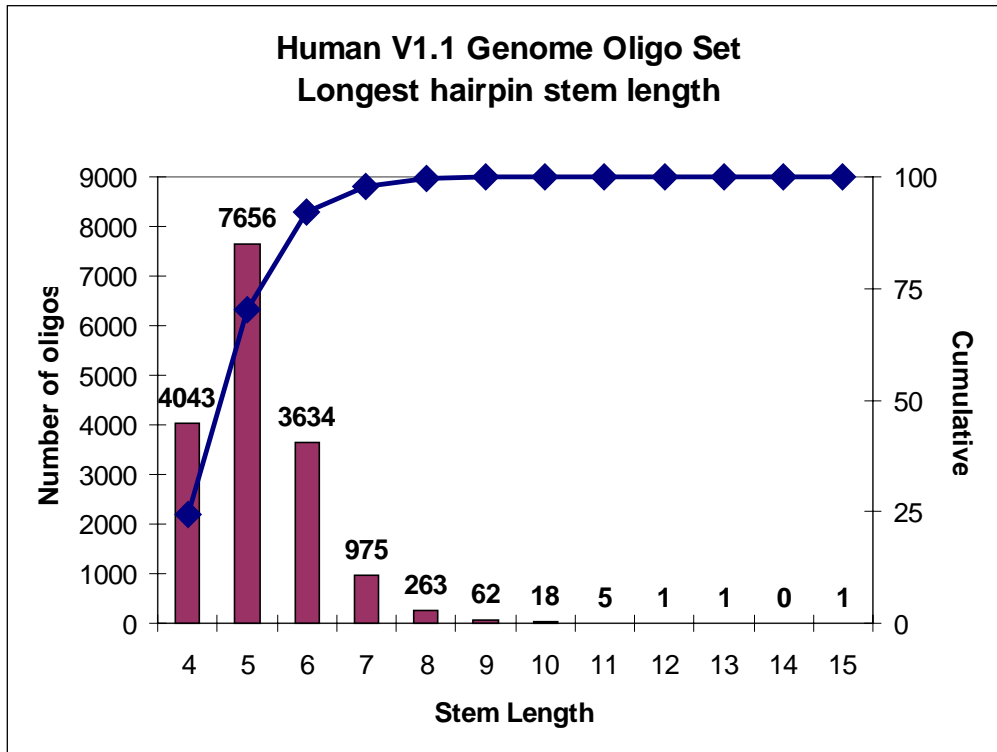Figure 4. Length of the Longest Hairpin Stem



**Human V1.1 Genome Oligo Set**
**Longest hairpin stem length**

Figure 5. Cross-Hybridization Identity



**Human V1.1 Genome Oligo Set**
**Cross-Hybridization Identity**